

Generating News Summaries at Indexing Time

Javier Parapar
IRLab, Computer Science Department
University of A Coruña
Fac. Informática, Campus de Elviña,
15071, A Coruña, SPAIN
javierparapar@udc.es

Abstract: This poster presents an efficiency oriented approach to the task of summary generation for operational news retrieval systems, where the summaries are appreciated by the users. This work shows that for this task the relevant sentence extraction techniques are suitable due to the compressibility of the generated summaries and the low computational costs associated. To minimize the costs of the summary construction in retrieval time we propose an efficient storage of the summaries as sentence offsets inside the documents. At indexing time the user query is not available to make the selection of the relevant sentence so the article's title was chosen to generate a title-biased summary, because of the high quality description of the news that the titles are. The sentence offsets were included in the direct file to just reconstruct the summaries in processing time from this information. This strategy gets a very high improvement in terms of retrieval time with a very low increment of the index size in comparison with query-biased summaries generated at retrieval time. As future work we will approach the evaluation of the summaries quality in base of the DUC measurements and the improvement of the relevance score formulas.

Keywords: Text summarisation, News systems, News indexing.